

UNCLASSIFIED

AD 403 435

*Reproduced
by the*

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

CATALOGED BY NSA
AS ADG NO. 403435

LINGUISTICS RESEARCH CENTER The University of Texas

FOURTEENTH QUARTERLY PROGRESS REPORT 1 AUGUST 1962-31 OCTOBER 1962

U.S. ARMY SIGNAL RESEARCH AND DEVELOPMENT LABORATORY MACHINE LANGUAGE TRANSLATION STUDY

CONTRACT NUMBER DA 36-039 SC 78911 • FILE NUMBER 18678-PM-59-91-91 (6909)

ASTIA

MAY 6 1963

**QUALIFIED REQUESTORS MAY OBTAIN
COPIES OF THIS REPORT FROM ASTIA**

FOURTEENTH QUARTERLY PROGRESS REPORT

1 AUGUST 1962 - 31 OCTOBER 1962

W. P. Lehmann, Director

E. D. Pendergraft, Associate Director

Prepared for

**U. S. Army Signal Research and Development Laboratory
Fort Monmouth, New Jersey
Contract No. DA 36-039 SC 78911
File No. 18678-PM-59-91-91 (6909)**

**LINGUISTICS RESEARCH CENTER
The University of Texas
Box 7247, University Station
Austin 12, Texas**

LRC 62-P14

November 1962

TABLE OF CONTENTS

| | | |
|-------|---|----|
| 1 | Purpose..... | 3 |
| 2 | Abstract..... | 5 |
| 3 | Publications, Lectures, Reports and Conferences.... | 7 |
| 4 | Research Summary..... | 11 |
| 5 | Progress in the Quarter..... | 13 |
| 5.1 | Systems Programming and Operations..... | 13 |
| 5.1.1 | Control Program..... | 14 |
| 5.1.2 | General Sort..... | 14 |
| 5.1.3 | Request Maintenance..... | 16 |
| 5.1.4 | Corpus Maintenance..... | 17 |
| 5.1.5 | Grammar Maintenance..... | 17 |
| 5.1.6 | Transfer Maintenance..... | 19 |
| 5.1.7 | Monolingual Recognition..... | 20 |
| 5.2 | Descriptive Linguistics..... | 21 |
| 5.2.1 | English..... | 21 |
| 5.2.2 | German..... | 24 |
| 5.2.3 | Russian..... | 26 |
| 5.2.4 | Chinese..... | 26 |
| 5.3 | Theoretical Linguistics..... | 27 |
| 5.3.1 | Theory of Formation Structures..... | 27 |
| 5.3.2 | Entropy in Stochastic Grammars..... | 28 |
| 6 | Conclusions..... | 29 |
| 7 | Planning for the Next Quarter..... | 31 |
| | REFERENCES..... | 33 |
| | PERSONNEL..... | 35 |

This report describes the progress of two projects. The first, entitled the Machine Language Translation Study, has been in progress since May 1959 under sponsorship of U.S. Army Signal Research and Development Laboratory. It has a long-range but primarily practical purpose: to implement automatic translation of languages by means of a large-scale, generalized computer system. The second, entitled the Development of a Linguistic Computer System, was initiated in September 1961 through a grant from the National Science Foundation. It has a broader purpose: to support basic research in linguistics by means of a computer system with generalized capabilities for language data processing and linguistic information processing. Those programs and portions of the work which primarily concern the development of a machine translation system are supported by the United States Army. The work directed primarily toward the linguistic computer system is supported by the National Science Foundation. These two projects complement each other in that programs prepared for each are applicable to, and needed in, the research of the other.

Progress is reported toward the implementation of a computer system to support mechanical translation and other linguistic processes which are potentially applicable to scientific documentation. The system contains three sections: one for control, a second for language data processing, and a third for linguistic information processing. The first is now operational, and the second will be so in the next quarter. The third is now being developed at the level of syntactic analysis.

Dr. W. P. Lehmann visited the machine translation research group at the University of California, Berkeley, California, on 1-2 August. He consulted with Dr. S. Lamb and other members of his group about their work in Russian and Chinese.

On 1-3 August Mr. E. Pendergraft visited the Centre de Traitement de l'Information Scientifique (CETIS) at the EURATOM Centre Commun de Recherche, Ispra, Italy. He conferred with Dr. Y. Lecerf about the work in mechanical translation under his direction, and with Dr. D. G. Hays who has taken a year's leave of absence from the RAND Corporation to pursue research in semantics as a guest at Ispra. Mr. Pendergraft talked at length with Dr. M. Detant concerning work on information retrieval in progress at the center, and with Dr. J. Verheyden about his work on syntax. On 2 August he presented the current work of the Linguistics Research Center to members of the EURATOM mechanical translation group.

Mr. Pendergraft went to the Centro di Cibernetica e di Attività Linguistiche at the University of Milan, Italy, on 4-7 August. He conferred with Prof. S. Ceccato, who directs research on automatic language translation at the center, and discussed the project's approach to semantics with Mr. E. Glasersfeld. Through the assistance of Mr. E. Marietti, he became familiar with some of the programming techniques being used at the center.

During the week of 27-31 August the Ninth International Congress of Linguists was held jointly by M.I.T. and Harvard University in Cambridge, Mass. Members of the Linguistics Research Center in attendance were Drs. W. P. Lehmann, W. Tosh, S. N. Werbow, W. Winter and Miss A. Yue. Papers were presented to sections of the congress by Dr. Lehmann ("Types of Sound Change", Section on Linguistic Change), Dr. Tosh ("Content Recognition and the Production of Synonymous Expressions", Section on Applications of Computers), and Dr. Winter ("Styles as Dialects", Section on Stylistics). All papers will be published under the title Proceedings of the Ninth International Congress of Linguistics. An opportunity for numerous conversations with linguists interested in different aspects of machine translation research was afforded by the congress. During the week Dr. Tosh and Dr. Werbow discussed the status of linguistic work under Contract Grant NSF G-19277 with Mr. R. See of the National Science Foundation.

On 24 September Dr. W. P. Lehmann and Dr. Tosh discussed our linguistic work briefly with Mr. E. Companys, who is Chargé d'études with the Bureau d'Études et de Liaison pour l'Enseignement du Français dans le monde, Paris.

The Linguistics Research Center was visited on 9-11 October by a group of specialists from Germany who are making a survey of mechanical translation research in this country as a basis for recommendations to their government concerning the initiation of a program of its own. Mr. R. F. Krollman, Director

of the Übersetzerdienst der Bundeswehr, the translating bureau of the German Armed Forces at Mannheim, was a member of the group. This agency employs some 300 professional translators and editors to translate all technical literature required by the Bundeswehr. Mr. D. Himburg represented the Military Research and Development Division of the German Ministry of Defense; he is responsible for sponsorship of basic and applied research in data processing, computation, and allied fields. Dr. H. Schnelle, Associate Professor at the University of Bonn, was also present. Because of his interests in cybernetics, information storage and retrieval, and mathematical linguistics, Dr. Schnelle holds a German defense research contract on preliminary research that will lead to mechanical translation. Mr. G. Beyer, Chief Interpreter on the staff of the German Military Representative to MC/NATO in Washington, D. C., represented the interests of the foreign language service of the German Federal Ministry of Defense in the United States. On 10 October Mr. Pendergraft presented our work in theoretical linguistics to the group. An explanation of our work in descriptive linguistics was given on the following day by Dr. Tosh. At that time Dr. Schnelle also outlined his own approach to mechanical translation. In conferences with Dr. W. P. Lehmann it was decided that Miss E. Hoffmann, a member of Mr. Krollman's staff, would join our descriptive linguistics group for six months to familiarize herself with our methods. Her work will be supported by the German government.

Mr. E. Pendergraft is conducting an advanced seminar on mechanical translation within the Linguistics Program of The University of Texas during the fall semester.

The research under these contracts is currently being conducted by three separate research groups whose responsibilities are as follows:

A theoretical linguistics group, consisting chiefly of mathematicians, has responsibility for development of the hypothesis underlying the work of the other two groups. An early result of the study was the conclusion that current theories of linguistic structure were not explicit enough to support applications in the field of mechanical translation. The first phase of the research was therefore concerned with attempts to explicate existant linguistic theories by means of formalization. This work has progressed to include formalized hypotheses for the syntactic and semantic relations which must be taken into account in translation. A general theory of translation may be based on this foundation; it has been completed in all important details. Present theoretical studies are designed to extend and further explore implications of the translation theory.

A systems group has responsibility for development of a generalized computer system based on the above theory, and for its operation in the performance of linguistic research on various languages. All essential features of the System have now been specified with sufficient precision for programming. It contains three main sections: one for control, a second for language data

processing, and a third for linguistic information processing, including three types of translation [1]. The first section is now operational, and the second will be so in the next quarter. The third section, which depends upon the second for its data, is at present under development at the level of syntactic analysis.

A descriptive linguistics group is engaged in testing the hypothesis by applying it to the German and English languages. Syntactic and semantic studies have been in progress for approximately three years; at present they are oriented to a specific corpus taken from Edward Ruechardt's book, Sichtbares und Unsichtbares Licht, and its English translation [2]. German and English dictionaries have also been based upon Der Sprach Brockhaus and Webster's New Collegiate Dictionary, respectively.

5 PROGRESS IN THE QUARTER

Research in the quarter included the following activities:

5.1 Systems Programming and Operations

The current objective in programming remains as stated in the previous report: to finish all programs needed to support lexical and syntactic monolingual recognition by the end of the year. Flow-charting and coding for this purpose have been completed, and all of the necessary programs are now being tested under careful monitoring to insure that this goal will be met.

Although program testing was performed primarily at the U. S. Army Electronic Proving Ground at Fort Huachuca, Arizona, toward the end of the quarter it became advantageous to move programs to the CEIR Computer Center in Houston. Better availability of computer time, together with greater mobility of the programming staff, made this arrangement more efficient for final checkout and assembly of sub-programs.

As grammar maintenance approached operational status, the operations section prepared for its increased responsibility by reviewing data handling and accounting procedures. The processing of grammatical data began near the end of the quarter as predicted, but these first uses of grammar maintenance programs are considered as program testing; the programs have not as yet been released for standard operational use.

The following progress was made in individual programs.

5.1.1 Control Program

The addition of new functions has caused the Control Program to become too large for the space allotted to it within the System; thus it became necessary to store infrequently used segments of the program on the Program Tape. The segments will now be called into core only when they are needed during System operation.

Two additional improvements were made in the Control section along with the above modification. Previously the Program Tape had been generated on the central computer and was available only during the processing mode of System operation [3, p. 3]. This tape may now be created by an off-line card-to-tape operation and will be accessible during both processing and checkout mode.

Most of the modification was completed in the quarter; it included the coding and testing of a bootstrap routine to initiate self-loading of the Control Program from the Program Tape and a routine to load other programs from the Program Tape.

5.1.2 General Sort

Testing of General Sort was completed early in the quarter. The program was immediately incorporated into the System, where its use in tests of other programs proved it to be accurate.

Experience with the program has already begun to indicate data situations for which the program may not be as efficient as specialized sorts. Every heuristic can be expected to be inefficient for some kinds of data, however, and the present one seems well suited to our needs. Study of the data situations which evidence slow sorting has suggested additional heuristics which may improve the generality of the program. These adjustments have been given low priority, since our objective at this phase of the study is not complete optimization but only optimization within the limits of good research procedure. General sorts in the System may be replaced by more specialized sorts at a later date when higher priority programming and more careful studies of sorting times have been completed.

General Sort is self-contained; it may be used either as a subroutine or as a complete program. Its data must be contained on an input tape with one sort item per record. General Sort may read fixed or variable length records, as specified. Two tapes are employed in the sort; the original input tape may be saved, in which case three tapes are required for sorting.

Three types of sort are available; these may be performed individually or in any combination for a given set of data. Adaptation of General Sort to a specific problem is achieved by means of a table, each entry of which specifies a type of sort to be performed and the conditions for sorting.

The sorts now available are:

- (1) Sort a word of each record according to the bit mask supplied with this entry. The word to be sorted is specified in this entry.
- (2) Sort a table of words in each record according to the bit mask supplied with this entry. The record word in which the table begins and the record word containing the table length are specified in this entry.
- (3) Sort a table of words in each record according to the bit mask supplied with this entry. The record word containing the beginning location of the table and the record word containing the table length are specified in this entry.

Additional sort types may be added to General Sort as desired; the above three have been found to be adequate for our present needs.

5.1.3 Request Maintenance

A relatively simple program to update, sort and verify certain features of Request Tapes has been found advantageous as an aid to systems operations. The program may be used to correct obvious errors that are discovered when data are being assembled for corpus, grammar or transfer maintenance thus avoiding more costly correction procedures in the System.

The program will be especially useful for the data of new languages, since these may be assembled on Request Tapes until studies have progressed sufficiently to merit the cost of their maintenance within the System. Our Chinese and Russian studies, for example, will initially have this status.

Design, flowcharting and coding for the program were completed; it should be operational early in the next quarter. The 416,000 cards containing German and English dictionaries will then be preprocessed by the program before these data are presented to the grammar maintenance process.

5.1.4 Corpus Maintenance

Final testing of the Corpus Revision and Corpus Display functions was completed, and the programs were instated on the Program Tape. Previously, corpus maintenance had been performed under control of the SOS or FAP monitoring systems.

Coding and testing were completed for the only remaining corpus maintenance function, Corpus Selection, which prepares requested samples of corpora as input for Lexical Recognition. Corpus Selection should be operational in the next quarter.

5.1.5 Grammar Maintenance

About half of the quarter was spent in testing individual segments of Grammar Revision. The ten segments of the

program were then combined and checked out as a unit with small samples of data. When these tests were completed, more comprehensive testing was performed with the complete German grammar based on Ruechardt's Sichtbares und Unsichtbares Licht [2].

Early experiences with real data demonstrated that a sort heuristic in the second segment of the program was unsatisfactory. After study a change in data handling procedures reduced the sorting time by almost one half, and further reductions are anticipated through the use of an additional heuristic. On the whole the programs perform well and should significantly reduce the cost of compiling and maintaining grammars.

The above experimentation with Grammar Revision led to the conclusion that some of the ten segments could be combined to yield a more compact program. As soon as the reassembled six-segment program has passed final tests, it will be used at Fort Huachuca to compile the German and English dictionaries.

Individual segments of Probability Revision had been almost completely tested at the end of the quarter. The program should be assembled and tested as a whole early in the coming quarter.

Those parts of Grammar Display needed for syntactic and semantic data were coded, tested and made available in the System. The results of Grammar Revision experiments were displayed with the program; it currently provides listings of

syntactic or semantic rules in two different sorts, either or both of which may be requested:

- (1) numerical sort on form number
- (2) alphabetical sort on all elements of the rule.

Both sorts are achieved with General Sort, using the first and second types of sorts mentioned above. The segments which will provide displays of lexical data and the mnemonic equivalents of syntactic or semantic variables have not yet been coded. These are not essential to current goals and will be added as time permits.

Testing of the individual segments of Input Grammar Selection was completed. The segments were combined and check-out begun on the whole program. Although the program was not operational at the end of the quarter, it is expected to be ready soon to prepare comprehensive test data for Monolingual Recognition.

5.1.6 Transfer Maintenance

Transfer maintenance programming continues to have low priority at present. The only activity in this area was completion of testing for the first of three segments of Interlingual Transfer Revision. This segment was then combined with the second, and testing of the pair was begun.

5.1.7 Monolingual Recognition

Testing of Lexical Monolingual Analysis and Lexical Monolingual Analysis Choice was completed; the two programs will be combined early in the next quarter so that their coordination may be checked with more complex data.

Test data were prepared and preliminary testing was started for Syntactic Monolingual Analysis. Considerable progress was made in debugging toward the end of the quarter. The program should be ready for checkout with more complex data early in the next quarter.

Preliminary testing of Syntactic Monolingual Analysis Choice was finished early in the quarter. Since the analysis program was not ready for the services of this routine, more complex data were prepared to test the choice heuristic. These experiments were still in progress at the end of the quarter.

Lexical and Syntactic Monolingual Analysis Display have been combined into a single program structure with two segments. The first segment gathers results of analysis and processes them into the order they will have in the display; the second produces the display itself. Testing of the individual segments is proceeding satisfactorily. Increasingly complex data are being given to the first segment, which is the more intricate of the two. The second segment now works for most test cases, and should soon be available for assembly with the first.

No additional work was done on Semantic Monolingual Recognition in the quarter.

5.2 Descriptive Linguistics

Pilot studies for Russian and Chinese were initiated during the quarter because of the increasing availability of structural information about these languages from other groups. With this addition, the linguistics group was organized into research sections for individual languages, and was renamed in an attempt to more accurately designate its responsibility. Progress will therefore be reported by language section.

5.2.1 English

During the tenth quarter a concerted effort was initiated to encode the basic German and English dictionaries for the translation system [4, p.17]. From that time the various parts of speech were encoded just as found in the dictionary references, or without certain specified affixes. Consequently any complex entries, such as noun compounds, derived adjectives, etc., were encoded as whole forms or as forms with only inflectional affixes removed.

Dictionary data are as a result not all in the final form which would be desired for translation. It would not be economical, for instance, to maintain noun compounds in unit form. We have accordingly continued with an analysis of the

internal morphological constituency of complex entries, as was first reported in the thirteenth quarter [1, p.40].

Internal analysis was completed in the quarter for verbs. Because nouns constitute by far the greatest portion of dictionary data, a considerable amount of work remains in this area; at the present time, internal analysis of nouns is approximately one-third complete.

In the area of syntax, a complete catalogue was made of the specific tagmemic patterns of sentences of Corpus 5 [2]. We now feel reasonably confident in our understanding of internal structures of the noun and adjective phrases and of the adverbials, some of which would be identified internally as noun or prepositional phrases. These elements were cataloged by recording the appropriate sequences of variables as they occur in clauses. All other elements, such as connectives or phrases which cannot as yet be classified, were recorded in the clause sequences as constants.

The first sentence of Corpus 5, for example, reads:

There is a beautiful expression for the arrival
in this world of a new member of the human family:
the infant "sees the light of day".

The clause sequence of this sentence was recorded as follows:

05 001

there VERB/SING NP/SING for NP/SING in NP/SING of
NP/SING of NP/SING: NP/SING "sees the light of day".

The interrupting quotation marks make it difficult to classify the expression "sees the light of day" consistently with other

data so far collected. The expression is entered as a sequence of constants to mark it for later analysis. In the recorded data, lower case letters represent constants. The remaining symbols designate syntactic variables.

Note that for the time being all prepositions were recorded as constants. Furthermore, we did not try to classify clause sequences at this phase of the analysis, but did so only after all clause data from Corpus 5 had been assembled. Clause rules were then written to describe the structures with the greatest possible generality within the limitations of the data. This technique will later be extended to include additional corpora by means of automatic syntactic analysis.

Structural diagrams were also prepared for all 1000 sentences of Corpus 5, except for a small residue of problematic occurrences including primarily those unique or near-unique structures which occur only once or a few times in the corpus. Since we have little else with which to compare these problematic structures, no attempt at general solutions shall be made without more data. A record is kept of such problems so that they may be reviewed as new circumstances dictate.

Collection of synonymy data has been concentrated into two activities [5]. The first is a study of synonymous substitutes for nominals occurring in Corpus 5. This study was done from the point of view of translation; the data consist of English nominals corresponding as equivalents to German

nominals in the original corpus. The second is a study of synonymous substitutes in general within the same corpus. It is not limited to specific syntactic categories like nominals, but may include as synonymous substitutes expressions of any category from the size of words to sentences, and without consideration of any language except English. To date, synonymous nominal expressions have been recorded for occurrences in the first 408 sentences of Corpus 5; the more general substitutes have been recorded for 616 sentences.

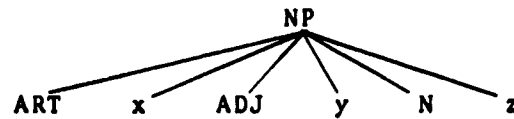
5.2.2 German

Additional work was done on the internal morphological constituency of German dictionary entries in studies roughly paralleling those already described for English. No internal analysis has as yet been done on German verbs. Adjective analysis is approximately one-third complete. Priority has been given to nouns, since this category constitutes the bulk of dictionary data. An accurate estimate of the status on noun analysis is not presently available; however, a working paper explaining details of German noun coding and the subsequent morphological analysis is in preparation [5]. It will be made generally available to interested individuals or research groups.

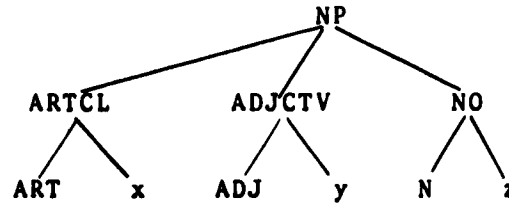
Three basic approaches to the organization of German noun phrase data have resulted from earlier studies. The

general form of these alternatives is shown below; constants functioning as inflectional suffixes are represented by x, y and z:

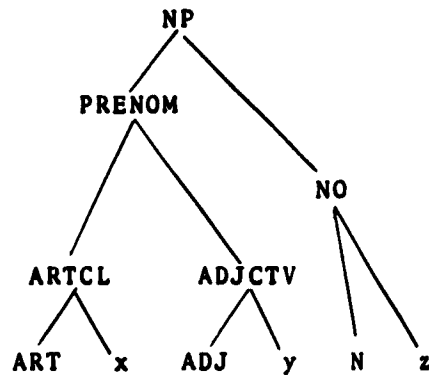
(1)



(2)



(3)



The alternatives will be evaluated with additional corpus data when automatic syntactic analysis becomes available.

A survey of clause components, similar to the English study described above, has been completed for the German version of Corpus 5. Structural diagrams of all sentences in the corpus were also prepared.

Compilation of German synonymy data also parallels the methods outlined above. Synonymous nominal expressions have been provided for all occurrences of nominals in German Corpus 5. Synonyms for adjectivals and other pre-nominal modifiers are being compiled at the present time.

5.2.3 Russian

Russian studies will concentrate first upon a structural description of the grosser elements of sentences in specific corpora. Three articles from Voprosy Ekonomiki, yielding a corpus of over 500 sentences, were selected for the pilot study [6]. The description was limited to structural relations obtaining between clauses and the constituents of individual clauses. The analysis of all 500 sentences was completed, and structural diagrams were prepared for all occurrences of clause structures. The description is currently being reviewed for consistency.

5.2.4 Chinese

Chinese studies were initiated with the same methodology as the Russian. For the first corpus we are using a text on language teaching which consists of 1357 sentences [7].

The clause analysis and all structural diagrams for the corpus have been completed. A review of resultant data is in progress.

5.3 Theoretical Linguistics

The responsibilities of the mathematics group were also broadened to include library research and the development of other sources of information by which we may support our efforts in descriptive linguistics. Mathematical research will continue within this wider framework.

The preparation of publications summarizing research results was emphasized in the quarter. Two new studies were also begun: one pertaining to the concept of entropy in stochastic formation structures, the other concerned with a model containing as "sub-structures" all phases of the translation process. These are both tentative orientations with a longer-range investigation of problems of optimization in linguistic description and translation. Since the entropy study has already reached completion, it is reported in somewhat greater detail below.

5.3.1 Theory of Formation Structures

The development of formation structure theory, along with its interpretations in syntactics and semantics, has

acquired such dimensions that the publication will be divided and published in smaller units. The first such working paper, entitled "Q-Collections and Concatenation", was completed, and will be distributed in the next quarter. The paper is concerned with basic properties of sequences of symbols occurring in the rules of formation structures [8, appendix]. Interpretations of the theory are illustrated by numerous linguistic examples.

The second paper in the series, concerned with the choice of axioms for formation structures and derived properties, is essentially complete, although theorems were added to it during the quarter. It is anticipated that this working paper may also be completed in the next quarter.

5.3.2 Entropy in Stochastic Grammars

The delimitation of various concepts of entropy in grammars for stochastic formation structures was undertaken in order to explore possible relevance to techniques of optimization. An investigation was first directed toward a quality evaluation of the stochastic grammar in terms of probability distributions over those sets of rules which define the membership of individual syntactic classes. The entropy computation may provide a basis for improving the grammar through reclassification. A method was also indicated for an entropy computation to evaluate the grammar as a whole. These results are being prepared as a working paper for general distribution.

It is anticipated that all programming for Corpus Maintenance, Grammar Maintenance, and Lexical and Syntactic Monolingual Recognition will be operational in the next quarter. The maintenance programs will probably be ready before the end of November, with completion of the recognition programs following about one month later, in mid-December.

These programs are essentially a compiler for syntactic recognition algorithms and for the data upon which such algorithms operate. They are designed primarily for purposes of experimentation with large stores of language data, though the principles upon which they are based may readily be adapted to a broader range of applications. This step has not been taken because it would be premature and wasteful. At present the most urgent need in translation research is more accurate knowledge of syntax.

We believe, therefore, that the capability of these programs will represent an important milestone in this project, and perhaps in the field as a whole, since it will appreciably shorten the time-scale for experimentation with syntactic analysis, and in consequence for development of applications. Our studies of Russian and Chinese have been undertaken to make better use of this capability, and to give it a broader test.

The systems programming section will concentrate upon completion of programs supporting syntactic recognition, and will then undertake the necessary indoctrination to make them available to the operations section. This will include preparation of necessary documentation; in particular, the users' manuals and appropriate flow diagrams will be developed for publication. As already mentioned, some features of the programs which do not interfere with immediate use will be added at convenient times during future programming schedules. This procedure is more efficient than rigidly defined scheduling; it will also be employed for those inevitable adjustments in programs which result from their first large-scale use.

As German and English data become available through the compilation of grammars, the descriptive linguistics group will give priority to error detection and correction in preparation for the first grammar revision cycle. For the time being, two grammars of each language will be maintained in the System: a smaller corpus-oriented one for experimentation, and one for comprehensive description of the language. Efforts will be made in the next quarter to clear up residual problems in German and English clause structures.

Studies of Russian and Chinese syntax will be continued in their present status: resultant data will be accumulated on Request tapes, but will not be maintained in the Language Data Processing Section.

Work in theoretical linguistics will be a continuation of the activities outlined above, with emphasis in the coming quarter again upon completion and distribution of research publications now in preparation. Planning will begin on specific goals for library research, and for maintenance of a glossary of the technical terms being used in our publications.

REFERENCES

- [1] Thirteenth Quarterly Progress Report for period
1 May 1962 - 31 July 1962.
- [2] E. Ruechardt, Sichtbares und Unsichtbares Licht,
Berlin, 1952.

_____, Light, Visible and Invisible, Ann Arbor,
1958.
- [3] R. W. Jonas, System Control, Programming Refer-
ence Manual, Linguistics Research Center, LRC
62-TP1, Austin, 1962.
- [4] Tenth Quarterly Progress Report for period
1 August 1961 - 31 October 1961.
- [5] G. R. Lewis, Report on German Noun Coding, Lin-
guistics Research Center, LRC 62-WDG1, Austin, 1962.
- [6] B. SHCHETININ, "EKONOMICHESKAJA POMOSHCH
SOCIALISTICHESKIKH STRAN MOLODYM NACIONAL'NYM
GOSIDARSTVAM", VOPROSY EKONOMIKI, 1960, vol. III,
no. 6, pp. 60-70.

B. SVETLICHNYI, "SOVETSKOE GRADOSTROITEL'STVO
NA SOVREMENNOM ETAPE", VOPROSY EKONOMIKI,
1960, vol. III, no. 8, pp. 48-60.

D. ALLAKHVERIAN, "O FUNKCIJAKH FINANSOV PRI
SOCIALIZME", VOPROSY EKONOMIKI, 1960, vol. III,
no. 11, pp. 45-55.

- [7] Yam Chung Ming, [Talks on the Rudiments of Language Teaching], Hong Kong, 1961.
- [8] Ninth Quarterly Progress Report for period 1 May 1961 - 31 July 1961.

PERSONNEL*

Administration:

Dr. W. P. Lehmann, Director (130)

Mr. E. D. Pendergraft, Associate Director (92)

Office:

Mrs. Z. Koprivnik, Secretary

Mr. A. Dominguez

Descriptive Linguistics:

Dr. L. W. Tosh, Chief (184)

Dr. E. Bach, Consultant

Dr. S. N. Werbow, Consultant (130)

Dr. W. Winter, Consultant (130)

Chinese:

Miss A. Yue

English:

Miss C. E. Barrett

Mrs. B. Efrat

Mr. I. Feigenbaum

Mr. C. Hayes

Mrs. M. S. Prince

German:

Mr. T. Baker

Mrs. D. E. Casey

Mrs. K. S. Heilbrunn

Mr. G. R. Lewis

Mrs. N. Orme-Johnson

Mrs. E. Sessions

Russian:

Dr. H. K. Ulatowska

Systems:

Mr. R. W. Jonas, Chief (184)

Operations:

Mrs. R. Ambagis (390)

Mr. C. A. Bramblett (110)

Mrs. M. Burkland (520)

Mr. J. Lo

Mr. O. H. Olson, Jr. (264)

Programming:

Mrs. B. C. Foster (184)

Mrs. M. Gordon

Mrs. M. L. Hagemeyer (184)

Mrs. O. Janca (520)

Mr. S. A. McGall (336)

Theoretical Linguistics:

Mr. W. B. Estes, Chief

Library:

Mr. J. Bunnag

Mathematics:

Miss J. Brady

Mr. J. M. Feeley

Mr. W. A. Holley

Mr. D. A. Senechalle

Mr. D. R. Stocks, Jr.

*The number of manhours worked in the quarter under contract
No. DA 36-039 SC 78911 are indicated in parentheses.

DISTRIBUTION LIST

| | |
|---|----|
| OASD (R&E), Rm. 3E1065 ATTN: Technical Library The Pentagon Washington 25, D. C. | 1 |
| Chief of Research and Development OCS, Department of the Army Washington 25, D. C. | 1 |
| Commanding General U. S. Army Materiel Command ATTN: R&D Directorate Washington 25, D. C. | 1 |
| Commanding General U. S. Army Electronics Command ATTN: AMSEL-AD Fort Monmouth, New Jersey | 3 |
| Commander, Armed Services Technical Information Agency ATTN: TISIA Arlington Hall Station Arlington 12, Virginia | 10 |
| Commanding General USA Combat Developments Command ATTN: CDCMR-E Fort Belvoir, Virginia | 1 |
| Commanding General U. S. Army Materiel Command ATTN: AMCRE-RP-PE, Attn Mr. A. Maklin Bldg T-17 Gravelly Point, Virginia | 1 |
| Commanding Officer USA Communication and Electronics Combat Development Agency Fort Huachuca, Arizona | 1 |
| Commanding General U. S. Army Electronics Research and Development Activity ATTN: Technical Library Fort Huachuca, Arizona | 1 |
| Commanding General U. S. Army Materiel Command Attn: MACRD-RS-PE, Mr. A. Maklin Bldg T-17 Gravelly Point, Virginia | 1 |

| | |
|--|---|
| Chief, U. S. Army Security Agency Arlington Hall Station Arlington 12, Virginia | 2 |
| Deputy President U. S. Army Security Agency Board Arlington Hall Station Arlington 12, Virginia | 1 |
| Director, U. S. Naval Research Laboratory ATTN: Code 2027 Washington 25, D. C. | 1 |
| Commanding Officer and Director U. S. Navy Electronics Laboratory San Diego 52, California | 1 |
| Aeronautical Systems Division ATTN: ASAPRL Wright-Patterson Air Force Base, Ohio | 1 |
| Air Force Cambridge Research Laboratories ATTN: CRZC L.G. Hanscom Field Bedford, Massachusetts | 1 |
| Air Force Cambridge Research Laboratories ATTN: CRXL-R L.G. Hanscom Field Bedford, Massachusetts | 1 |
| HQ, Electronic Systems Division ATTN: ESAT L.G. Hanscom Field Bedford, Massachusetts | 1 |
| Rome Air Development Center ATTN: RAALD Griffiss Air Force Base, New York | 1 |
| AFSC Scientific/Technical Liaison Office U. S. Naval Air Development Center Johnsville, Pennsylvania | 1 |
| Commanding Officer U. S. Army Electronics Materiel Support Agency ATTN: SELMS-ADJ Fort Monmouth, New Jersey | 1 |

| | |
|---|---|
| Director, Fort Monmouth Office USA Communication and Electronics Combat Development Agency Fort Monmouth, New Jersey | 1 |
| Corps of Engineers Liaison Office U. S. Army Electronics Research & Development Laboratory Fort Monmouth, New Jersey | 1 |
| Marine Corps Liaison Office U. S. Army Electronics Research & Development Laboratory Fort Monmouth, New Jersey | 1 |
| AFSC Scientific/Technical Liaison Office U. S. Army Electronics Research & Development Laboratory Fort Monmouth, New Jersey | 1 |
| Commanding Officer U. S. Army Electronics Research & Development Laboratory ATTN: Logistics Division Fort Monmouth, New Jersey For D.D. Jacoby, SELRA/NPE | 3 |
| Commanding Officer U. S. Army Electronics Research & Development Laboratory ATTN: Director of Research Fort Monmouth, New Jersey | 1 |
| Commanding Officer U. S. Army Electronics Research & Development Laboratory ATTN: Technical Documents Center Fort Monmouth, New Jersey | 1 |
| Commanding Officer U. S. Army Electronics Research & Development Laboratory ATTN: SELRA/NPE Fort Monmouth, New Jersey | 1 |
| Commanding Officer U. S. Army Electronics Research & Development Laboratory ATTN: Technical Information Division Fort Monmouth, New Jersey | 3 |
| Commanding Officer U. S. Army Electronics Research & Development Laboratory ATTN: J. Benson, SELRA/XC Fort Monmouth, New Jersey | 1 |

| | |
|---|--------|
| Commanding Officer U. S. Army Electronics Research & Development Laboratory ATTN: SELRA /X Fort Monmouth, New Jersey | 1 |
| Ramo-Wooldridge Corporation Los Angeles, California ATTN: Information Systems Dept. | 1 |
| Carnegie Institute of Technology Schenley Park Pittsburgh, Pennsylvania ATTN: Dr. Alan Perlis Dr. Allen Newell | 1 1 |
| R.C.A. Data Systems Division 4922 Fairmont Avenue Bethesda 14, Maryland ATTN: Dr. Jack Minker | 1 |
| Commanding Officer U. S. Army Signal Corps School Officers Dept. Fort Monmouth, New Jersey ATTN: ADPS Committee | 1 |
| Commanding Officer U. S. Army Signal Center & School Dept. Specialist Training Fort Monmouth, N. J. ATTN: Information Processing Branch | 1 |
| Commanding Officer U. S. Army Signal Engineering Agency Washington 25, D. C. ATTN: Computer Systems Division | 1 |
| Lincoln Laboratories Lexington 72, Massachusetts ATTN: Army Liaison Officer | 1 |
| National Bureau of Standards Washington 25, D. C. ATTN: Mrs. Ida Rhodes Mechanical Translation Project ATTN: Mr. Russell Kirsch Data Processing Systems Division | 1 1 |
| National Science Foundation Washington 25, D. C. ATTN: Mr. Richard See Documentation Research Program | 1 |

Director, National Security Agency
Fort George G. Meade, Maryland
ATTN: C-3141 (Rm 2C087)
Librarian

2

Professor E. deGrollier
Institute National Des Techniques
De La Documentation
Paris, France

1

International Electric Corporation
Box 285
Attn: J. Harlow
Paramus, New Jersey

1

| | |
|--|--|
| <p>The University of Texas, Austin, Texas MACHINE LANGUAGE TRANSLATION STUDY W. P. Lehmann Chief Investigator</p> <p>Fourteenth Quarterly Progress Report, 1 August 1962 to 31 October 1962, 37 pp. Signal Corps Contract DA 36-039 SC 78911 File No. 18678-PM-59-91-91 (6909) Unclassified Report.</p> <p>Progress is reported toward the implementation of a computer system to support mechanical translation and other linguistic processes which are potentially applicable to scientific documentation. The system contains three sections: one for control, a second for language data processing, and a third for linguistic information processing. The first is now operational, and the second will be so in the next quarter. The third is now being developed at the level of syntactic analysis.</p> | <p>Unclassified</p> <ol style="list-style-type: none"> 1. Machine Language Translation Study 2. Signal Corps Contract DA 36-039 SC 78911 |
| <p>The University of Texas, Austin, Texas MACHINE LANGUAGE TRANSLATION STUDY W. P. Lehmann Chief Investigator</p> <p>Fourteenth Quarterly Progress Report, 1 August 1962 to 31 October 1962, 37 pp. Signal Corps Contract DA 36-039 SC 78911 File No. 18678-PM-59-91-91 (6909) Unclassified Report.</p> <p>Progress is reported toward the implementation of a computer system to support mechanical translation and other linguistic processes which are potentially applicable to scientific documentation. The system contains three sections: one for control, a second for language data processing, and a third for linguistic information processing. The first is now operational, and the second will be so in the next quarter. The third is now being developed at the level of syntactic analysis.</p> | <p>Unclassified</p> <ol style="list-style-type: none"> 1. Machine Language Translation Study 2. Signal Corps Contract DA 36-039 SC 78911 |
| <p>The University of Texas, Austin, Texas MACHINE LANGUAGE TRANSLATION STUDY W. P. Lehmann Chief Investigator</p> <p>Fourteenth Quarterly Progress Report, 1 August 1962 to 31 October 1962, 37 pp. Signal Corps Contract DA 36-039 SC 78911 File No. 18678-PM-59-91-91 (6909) Unclassified Report.</p> <p>Progress is reported toward the implementation of a computer system to support mechanical translation and other linguistic processes which are potentially applicable to scientific documentation. The system contains three sections: one for control, a second for language data processing, and a third for linguistic information processing. The first is now operational, and the second will be so in the next quarter. The third is now being developed at the level of syntactic analysis.</p> | <p>Unclassified</p> <ol style="list-style-type: none"> 1. Machine Language Translation Study 2. Signal Corps Contract DA 36-039 SC 78911 |
| <p>The University of Texas, Austin, Texas MACHINE LANGUAGE TRANSLATION STUDY W. P. Lehmann Chief Investigator</p> <p>Fourteenth Quarterly Progress Report, 1 August 1962 to 31 October 1962, 37 pp. Signal Corps Contract DA 36-039 SC 78911 File No. 18678-PM-59-91-91 (6909) Unclassified Report.</p> <p>Progress is reported toward the implementation of a computer system to support mechanical translation and other linguistic processes which are potentially applicable to scientific documentation. The system contains three sections: one for control, a second for language data processing, and a third for linguistic information processing. The first is now operational, and the second will be so in the next quarter. The third is now being developed at the level of syntactic analysis.</p> | <p>Unclassified</p> <ol style="list-style-type: none"> 1. Machine Language Translation Study 2. Signal Corps Contract DA 36-039 SC 78911 |